融合隐含信任度和项目关联度的矩阵分解推荐算法 *

李　全，许新华，刘兴红，林　松

(湖北师范大学 教育信息与技术学院, 湖北 黄石 435002)

摘　要：随着社交网络的发展，融合社交信息的推荐系统在一定程度上解决了协同过滤推荐系统的冷启动和数据稀疏等问题，但是在信任数据稀疏情况下，仍会造成推荐精度降低等问题。为此，提出了一种融合隐含信任度和项目关联度的矩阵分解推荐算法。首先，利用矩阵分解模型将信任数据进行分解，得到用户的潜在被信任矩阵，在此基础上引入用户的影响力，从而提出了基于隐含信任度的推荐模型；然后，为了更好的利用项目间的关联信息，反映项目间的有向性，提出了基于项目关联度的推荐模型；最后，综合两种推荐模型并构建了一种推荐算法 TCRMF。实验结果表明，所提算法在评分数据和信任数据稀疏的情况下仍然可以有效地提高推荐算法的精度，具有良好的应用前景。

关键词：推荐系统；协同过滤；社交网络；隐含信任度；项目关联度；矩阵分解

中图分类号：TP301.6　　doi: 10.19734/j.issn.1001-3695.2018.07.0530

Matrix factorization recommendation algorithm combing implicit trust and item correlation

Li Quan, Xu Xinhua, Liu Xinghong, Lin Song

(*Collegeof Educational Information & Technology*，*Hubei Normal University*, *Huangshi Hubei* 435002, *China*)

Abstract: With the development of social network, recommendation system fusing social information sovled the problems of cold starting and sparse rating datas of collaborative filtering recommendation system to some extend. Therefore, it still caused problems of making recommendation accuracy decline in the case of the sparse trust datas. Thus, this paper proposed matrix factorization recommendation algorithm combing implicit trust and item correlation. Firstly, it decomposed the trust datas by the matrix factorization model, and obtained the implicit trusted matrix of users. It introduced the influence of users on this basis, and proposed the recommendation model based on implicit trust. Secondly, it proposed the recommendation model based on item correlation in order to make better use of correlation information between items and reflect on orientation between items. Finally, it combined the two kinds of recommendation models, and proposed the TCRMF algorithm. The experimental results show that the proposed algorithm still improves the accuracy of recommendation algorithm effectively in the conditions that score and trust datas are sparse, and has a good application prospect.

Key words: recommendation system; collaborative filtering; social network; implicit trust; item correlation; matrix factorization

0　引言

随着网路技术的不断发展，互联网上的资源与信息迅速的膨胀，如何从海量的数据中获取有效的信息，对于普通用户来说是一个巨大的挑战。解决这类问题最有效的方法就是个性化推荐系统。该系统通过分析用户的喜好，为用户推荐其感兴趣的信息[1]。现有的推荐方法一般可以分为三类：基于内容的推荐方法[2]、基于协同过滤推荐方法[3]和混合推荐方法[4]。在这些推荐方法中，基于协同过滤的推荐是应用最广泛的推荐方法。该方法主要包括基于用户(user-based)协同过滤、基于项目(item-based)协同过滤和基于矩阵分解(matrix factorization, MF)[5, 6]协同过滤等。但它们通常面临着冷启动、数据稀疏、算法可扩展性差等问题[7]。

近年来，随着社交网络的发展，用户之间的信任关系作为社交网络中一个重要的信息，为解决推荐系统中数据稀疏等问题提供了依据[8]。基于信任的推荐系统在一定程度上缓解了评分数据稀疏等问题，但是也带来了一些新的问题。目前，大多数基于社交网络的信任度是从用户评分数据或者社交信任图数据的角度进行度量。然而，在信任数据稀疏情况下，会造成推荐精度降低等问题。因此，如何在信任数据稀疏的情况下有效地挖掘用户之间的信任关系，度量用户之间的信任值成为影响社会化推荐算法进一步提高的重要问题。同时在人们的生活中，除了用户之间的社会关系外，物品之间也存在一定的关联关系。物品之间这种关联关系也会影响到用户对物品的选择。因此，如何获取物品之间关联关系并将其应用到推荐系统中是当前研究的难点问题之一。

为了进一步提高推荐算法的性能。本文提出了一种融合隐含信任度和项目关联度的矩阵分解算法。首先，利用矩阵分解模型将信任数据进行分解，得到用户的潜在被信任矩阵。接着，在该矩阵的基础上引入用户的影响力，并以此为基础，提出了局部隐含信任度和全局隐含信任度的信任度量方法和相应的推荐模型。然后，为了更好地反映项目间的关联关系

收稿日期：2018-07-23；**修回日期：**2018-09-26　**基金项目：**湖北省教育厅科技项目（B2018150）；国家档案局科技计划项目（2016-x-51）；湖北省教育科学“十二五”规划项目（2011B130）；湖北省高等学校优秀中青年科技创新团队计划项目（T201515）

作者简介：李全（1982-），男，湖北黄陂人，讲师，硕士，主要研究方向为机器学习、数据挖掘等（56322268@qq.com）；许新华（1968-），男，湖北孝感人，教授，硕士，主要研究方向为数据库、数据挖掘；刘兴红（1969-），女，湖北蕲春人，教授，硕士，主要研究方向为大数据、教育信息化；林松（1978-），男，湖北黄石人，讲师，硕士，主要研究方向为搜索引擎、数据挖掘.

有向性，提出了一种基于项目相关度和流行度的推荐模型。最后，综合考虑用户隐含信任度和项目关联度的推荐模型并提出相应的推荐算法。实验结果表明，本文提出的算法与其他社会化推荐算法相比，在均方根误差等方面都获得更加精确的推荐结果。

1　相关工作

典型的基于协同过滤推荐算法通常面临冷启动、数据稀疏等问题。随着社交网络的快速发展，社交信息为传统的协同过滤算法引入的新的数据源。因此，越来越多的研究者研究如何利用社交信息来提升推荐系统的质量。Massa 等人[9]首次提出基于信任的推荐系统，他们利用用户之间的相似性表示用户之间的信任值，实验结果表示基于信任的推荐系统有利于提高推荐系统的质量。王兴茂等人[10]提出了一种基于一跳信任模型的协同过滤推荐算法，他们在用户评分信息基础上定义用户直接和间接社会信任属性，进而计算用户之间的信任度，基于此模型设计的推荐算法提高了推荐准确度。赵海燕等人[11]提出了基于信任传播的概率矩阵分解算法，他们在信任图模型基础上定义了基于入度的信任度量和基于出入度的信任度量等三种信任度度量方法，并将它们融入到基于概率矩阵分解的社交推荐模型中，实验结果表明基于入度的信任度量方法的推荐效果较好。陈婷等人[12]提出了社交网络环境下基于信任的推荐算法，他们利用信任的传播性质对信任关系进行建模，在用户评分数据的基础上综合考虑相似度和信任度来构建用户间的偏好关系，验证了算法的高效性。郭磊等人[13]利用对评分矩阵进行概率矩阵分解得到的潜在向量计算用户与好友之间的相似性，提高了算法的精度。但是以上这些推荐算法不足是当用户评分数据或者社交信任图数据稀疏的情况下，根据它们计算得到的用户之间的信任度不能准确的反映用户与好友之间的相似关系，从而对推荐算法的精度会产生不利的影响。

除了用户之间的社会关系外，物品之间关联关系同样为协同过滤算法引入了新的数据源提供了依据。因此，研究如何利用物品之间的关联关系来提升推荐系统的精度是当前研究的重点问题。Wu 等人[14]通过标签信息来构造用户和物品的相似性关系，并将两者融入 PMF 中提出基于近邻的概率矩阵分解模型。杨兴耀等人[15]通过皮尔逊相关模型作为用户和物品的相似性度量模型，并通过 Jaccard 系数对相似性进行优化，提出一种综合用户和项目预测的协同过滤模型。于美琪等人[16]分析 Hownet 等语义词典中同义词组成的树状结构，通过计算物品之间的距离、深度和密度等语义距离，对物品进行基于语义相似的聚类，提出一种融合社会网络和物品特征的移动应用推荐。孙金刚等人[17]利用传统相似性计算方法得到物品之间的评分相似性，同时结合物品属性，计算物品的属性相似性，通过加权因子得到物品的最终相似性，提出一种基于物品属性和云填充的推荐算法。然而以上这些推荐算法不足是物品之间的相似关系都是无向的，根据它们计算得到的物品之间的相关度无法很好的反映物品之间真实的关联关系，从而对推荐算法的精度也会产生不利的影响。

针对以上两点问题，本文提出了一种融合隐含信任度和项目关联度的矩阵分解推荐算法，该算法的优势在于：

a)利用矩阵分解模型将信任数据进行分解，得到用户的潜在被信任矩阵，并在此基础上引入用户的影响力，提出了局部和全局隐含信任度的度量方法和相应的推荐模型，从而在一定程度上解决了信任数据稀疏情况下推荐精度降低等问题。

b)在项目皮尔森相似度的基础上加入惩罚系数，通过项目相关度和流行度分别从局部和全局的角度提出了项目间关联关系的度量方法和相应的推荐模型，从而能更好地反映项目间关联关系的有向性。

c)提出了一种综合考虑用户隐含信任度和项目关联度的推荐算法，从社交网络数据和项目关联数据两方面，进一步提高推荐算法的精度。

2　矩阵分解模型

2.1　问题描述

在推荐系统中含有由 n 个用户组成的集合 $U=\{u_1,u_2,\ldots,u_n\}$ 和 m 个项目组成的集合 $I=\{i_1,i_2,\ldots,i_m\}$，以及用户对项目的评分矩阵 $R=[r_{ui}]_{n\times m}$。如果用户 u 对项目 i 进行了评分，则评分矩阵中的元素 r_{ui} 表示相应的评分值；否则，$r_{ui}=0$ 表示用户 u 未对项目 i 进行评分。在社交网络中，每一个用户都有 N_u 个好友。用户的社交关系矩阵 $T=[t_{u,v}]_{n\times n}$。如果用户 u 和用户 v 具有社交关系，则社交关系矩阵中的元素 $t_{u,v}$ 表示用户间的信任强度，否则，$t_{u,v}=0$ 表示用户 u 和用户 v 没有社交关系。

2.2　低秩矩阵分解模型

为了预测推荐系统中用户-项目评分矩阵 $R_{n\times m}$ 的缺失值，矩阵分解模型通过降维的方法将高阶的评分矩阵 $R_{n\times m}$ 分解为用户 U 和项目 V 两个低维矩阵，如式(1)所示。

$$R\approx U^TV \tag{1}$$

其中：$U=[U_1,U_2,U_3,\ldots,U_n]\in R^{k\times n}$ 表示用户特征矩阵，U_i 表示用户 i 的 k 维潜在特征向量。$V=[V_1,V_2,V_3,\ldots,V_m]\in R^{k\times m}$ 表示项目特征矩阵，V_j 表示项目 j 的 k 维潜在特征向量。用户 i 对项目 j 的预测评分可表示为 $\hat{R}_{ij}=U_i^TV_j$。将用户原始评分与预测评分之间的误差的平方作为损失函数，如式(2)所示，通过最小化损失函数得到矩阵 U 和矩阵 V。

$$f(R,U,V)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}I_{ij}(R_{ij}-U_i^TV_j)^2 \tag{2}$$

其中：I_{ij} 是一个指示函数，当 $I_{ij}=1$ 表示用户 i 对项目 j 打过评分，否则 $I_{ij}=0$。为了防止过拟合，在式(2)的基础上添加两个正则化项，以对参数进行约束，如式(3)所示。

$$f(R,U,V)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}I_{ij}(R_{ij}-U_i^TV_j)^2+\frac{\lambda_U}{2}\|U\|_F^2+\frac{\lambda_V}{2}\|V\|_F^2 \tag{3}$$

其中：$\lambda_U>0$ 和 $\lambda_V>0$ 是正则化参数，$\|U\|_F$ 和 $\|V\|_F$ 表示 Frobenius 范数，采用随机梯度下降法得到用户特征矩阵 U 和项目特征矩阵 V，并可以准确的预测原评分矩阵 $R_{n\times m}$ 的缺失值。

3　融合隐含信任度和项目关联度推荐算法

3.1　基于隐含信任度推荐模型

在用户评分数据或者社交信任图数据稀疏的情况下，根据它们计算得到的用户之间的信任度不能准确的反映用户与好友之间的相似关系，会对推荐算法的精度产生不利的影响。为此，本节综合考虑了局部隐含信任度和全局隐含信任度的度量方法，并提出了综合隐含信任度的推荐模型。

3.1.1 局部隐含信任度

由文献[18]分析可知，已知用户的信任关系矩阵 S，通过低秩矩阵分解模型对该矩阵进行分解，得到用户的信任者特征矩阵 B 和被信任者特征矩阵 W。如式(4)所示。

$$S\approx B^TW \tag{4}$$

其中：$B=[B_1,B_2,B_3,\ldots,B_n]\in R^{d\times n}$ 表示用户的信任者特征矩阵，B_i 表示用户 i 作为信任者的 d 维潜在特征向量。

$W=[W_1,W_2,W_3,...,W_n]\in R^{d\times n}$ 表示用户的被信任者特征矩阵，W_j 表示用户 j 作为被信任者的 d 维潜在特征向量。用户 i 与用户 j 的预测信任强度可表示为 $\hat{S}_{ij}=B_i^TW_j$。将用户之间原始信任强度与预测信任强度之间的误差的平方作为损失函数，如式(5)所示，通过最小化损失函数得到矩阵 B 和矩阵 W。

$$f_1(S,B,W)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}I_{ij}(S_{ij}-B_i^TW_j)^2 \quad (5)$$

其中：I_{ij} 是一个指示函数，当 $I_{ij}=1$ 表示用户 i 与用户 j 存在信任关系，否则 $I_{ij}=0$。为了防止过拟合，在式(5)的基础上添加两个正则化项，以对参数进行约束，如式(6)所示。

$$f_1(S,B,W)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}I_{ij}(S_{ij}-B_i^TW_j)^2+\frac{\lambda_B}{2}\|B\|_F^2+\frac{\lambda_W}{2}\|W\|_F^2 \quad (6)$$

其中 $\lambda_B>0$ 和 $\lambda_W>0$ 是正则化参数，$\|B\|_F$ 和 $\|W\|_F$ 表示 Frobenius 范数。采用随机梯度下降法优化上式，B、W 的求导公式如式(7)和(8)所示。

$$\frac{\partial f_1}{\partial B_i}=\sum_{j=1}^{n}I_{ij}(B_i^TW_j-S_{ij})W_j+\lambda_B B_i \quad (7)$$

$$\frac{\partial f_1}{\partial W_j}=\sum_{i=1}^{n}I_{ij}(B_i^TW_j-S_{ij})B_i+\lambda_W W_j \quad (8)$$

通过矩阵分解算法将每一个用户映射为信任者的特征向量 B_i 和被信任者的特征向量 W_j。由文献[19]分析可知，在社交网络中，用户的影响力越大，那么该用户就越容易受到其他用户的信任。因此通过用户的被信任者的特征向量引入用户的影响力，如式(9)所示。

$$Influence(i)=\sqrt{W_i^TW_i} \quad (9)$$

其中：W_i 为用户 i 的被信任者特征向量。用户之间的信任关系是非对称的，例如你对好友的信任程度与好友对你的信任程度是不同的。所以为了准确地度量用户之间不同的信任关系，引入局部隐含信任度，如式(10)所示。

$$S_{local}(i,j)=\frac{Influcence(j)}{\sum_{k\in N_u}Influcence(k)} \quad (10)$$

其中：$Influcence(j)$ 表示好友 j 影响力，N_u 表示用户 i 的好友集合。好友 j 的影响力越大，该用户越容受到用户 i 的信任，即影响力大的好友对用户的作用越大。

3.1.2 全局隐含信任度

由文献[12]分析可知，全局信任度表示用户在社会网络中的的信任参考了其他所有节点对其的综合评价，即用户在整个社交网络中的可靠性和影响力。为了从全局角度度量用户之间的信任关系，引入全局隐含信任度，如式(11)所示。

$$S_{global}(i)=\frac{\sqrt{\left(Influcence(i)-Mean(\sum_{k=1}^{n}Influcence(k))\right)^2}}{Max(\sum_{k=1}^{n}Influcence(k))-Min(\sum_{k=1}^{n}Influcence(k))} \quad (11)$$

其中：$Influcence(i)$ 表示用户 i 的影响力；$Mean(\sum_{k=1}^{n}Influcence(k))$ 表示社交网络中所有用户影响力的均值；$Max(\sum_{i=k}^{n}Influcence(k))$ 表示影响力的最大值；$Min(\sum_{k=1}^{n}Influcence(k))$ 表示影响力的最小值。在社交网络中，用户 i 的影响力越大，该用户在社交网络中的全局信任度越大，越会受到更多人的信任。

3.1.3 综合隐含信任度推荐模型

在信任数据稀疏的情况下，根据局部隐含信任度和全局隐含信任度计算得到的用户之间的信任度能更加准确的反映用户与好友之间的相似关系。因此，综合两种隐含信任度并提出了相应的推荐模型，该模型如式(12)所示。

$$f_2(R,U,V)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}I_{ij}(R_{ij}-U_i^TV_j)^2+\frac{\lambda_U}{2}\|U\|_F^2+\frac{\lambda_V}{2}\|V\|_F^2+ \frac{\alpha}{2}\sum_{i=1}^{n}\sum_{j\in S_i^+}S_{local}(\mathrm{i,j})\|U_i-U_j\|_F^2+\frac{1-\alpha}{2}\sum_{i=1}^{n}\sum_{j\in S_i^+}S_{global}(\mathrm{j})\|U_i-U_j\|_F^2 \quad (12)$$

其中：S_i^+ 表示用户 i 的社会好友关系集合。参数 α 控制用户的局部隐含信任度和全局隐含信任度对推荐算法的影响。

3.2 基于项目关联度推荐模型

项目之间的关联关系在推荐系统中是影响用户决策的重要因素之一。本文从项目的相关度和流行度的角度对项目进行建模。

3.2.1 项目相关度

通常度量项目相似性的皮尔森相似度或者余弦相似度具有对称性，即项目 p 和项目 q 的相似性 $Sim(p,q)$ 等于 $Sim(q,p)$ [20]。然而在实际生活中，为考虑项目之间相互影响的差异性，项目 p 和项目 q 相似性与项目 q 和项目 p 的相似性是不同的。因此为了准确的描述项目间的关联关系，在不考虑其他项目相关信息的情况下，通过对皮尔森相似性加入惩罚系数，项目 p 对项目 q 的相关度如式(13)所示。

$$W_{associ}(p,q)=\frac{|N_p\cap N_q|}{|N_p|^{1-\gamma}|N_q|^{\gamma}}Sim(p,q) \quad (13)$$

其中：$|N_p|$ 表示项目 p 在推荐系统中对项目 p 进行评分的用户数，$|N_q|$ 表示项目 q 在推荐系统中对项目 q 进行评分的用户数。参数 γ 为惩罚因子，通常取值范围为 $\gamma\in(0.5,1)$，实现项目之间关联关系的有向性。$Sim(p,q)$ 表示项目间的皮尔森相似度。

3.2.2 项目流行度

项目流行度是从全局角度表示某项目在所有项目集中流行程度。推荐系统中项目的流行度一般具有较大的差异性，从项目的流行度可以推断用户对未评分项目感兴趣的可能性。项目的流行度越高，用户对未评价项目感兴趣的可能性越大。由文献[21]分析可知，项目的流行度与浏览过该项目的用户数量成正比，并且在计算每个项目的流行度时进行向量标准化，如式(14)所示。

$$W_{pop}(p)=\frac{|N_p|}{\sqrt{\sum_{p\in I}|N_p|^2}} \quad (14)$$

其中：$|N_p|$ 表示对项目 p 进行评分的用户数，I 表示所有项目的集合。

3.2.3 综合关联度推荐模型

项目相关度从局部角度获取项目间的有向关联关系，项目流行度从全局的角度对该关系进行修正，以更加准确的反映项目间关联度。因此，提出了一种综合关联度的推荐模型，该模型如式(15)所示。

$$f_2(R,U,V)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}I_{ij}(R_{ij}-U_i^TV_j)^2+\frac{\lambda_U}{2}\|U\|_F^2+\frac{\lambda_V}{2}\|V\|_F^2+ \frac{\beta}{2}\sum_{i=1}^{n}\sum_{j\in K_i^+}W_{associ}(\mathrm{i,j})\|V_i-V_j\|_F^2+\frac{1-\beta}{2}\sum_{i=1}^{n}\sum_{j\in K_i^+}W_{pop}(\mathrm{i})\|V_i-V_j\|_F^2 \quad (15)$$

其中：K_i^+ 表示项目 i 紧密关联的项目集合。参数 β 控制着项目的相关度和项目流行度对推荐算法的影响。

3.3 融合隐含信任度和项目关联度推荐模型

基于隐含信任度推荐模型从社交网络角度分析用户间的信任度来提升推荐系统精度，而基于项目关联度推荐模型从项目间关联关系角度分析项目间的关联度来提升推荐系统精度。因此，综合考虑两种方法，提出融合隐含信任度和项目关联度的推荐模型，该模型如式(16)所示。

$$f_2(R,U,V)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}I_{ij}(R_{ij}-U_i^TV_j)^2+\frac{\lambda_U}{2}\|U\|_F^2+\frac{\lambda_V}{2}\|V\|_F^2+$$
$$\frac{\sigma}{2}\left(\alpha\sum_{i=1}^{n}\sum_{j\in S_i^+}S_{local}(\mathrm{i,j})\|U_i-U_j\|_F^2+(1-\alpha)\sum_{i=1}^{n}\sum_{j\in S_i^+}S_{global}(\mathrm{j})\|U_i-U_j\|_F^2\right)+ \quad (16)$$
$$\frac{\rho}{2}\left(\beta\sum_{i=1}^{n}\sum_{j\in K_i^+}W_{associ}(\mathrm{i,j})\|V_i-V_j\|_F^2+(1-\beta)\sum_{i=1}^{n}\sum_{j\in K_i^+}W_{pop}(\mathrm{i})\|V_i-V_j\|_F^2\right)$$

其中：参数 σ 表示用户的社会关系对推荐算法的影响。参数 ρ 表示项目的关联关系对推荐算法的影响。采用随机梯度下降法优化上式，U、V 的求导公式如式(17)和(18)所示。

$$\frac{\partial f_2}{\partial U_i}=\sum_{j=1}^{m}I_{ij}(U_i^TV_j-R_{ij})V_j+\lambda_U U_i+$$
$$\sigma\left(\alpha\sum_{j\in S_i^+}S_{local}(\mathrm{i,j})(U_i-U_j)+(1-\alpha)\sum_{j\in S_i^+}S_{global}(\mathrm{j})(U_i-U_j)\right) \quad (17)$$

$$\frac{\partial f_2}{\partial V_j}=\sum_{i=1}^{n}I_{ij}(U_i^TV_j-R_{ij})U_i+\lambda_V V_j+$$
$$\rho\left(\beta\sum_{i=1}^{n}W_{associ}(\mathrm{i,j})(V_i-V_j)+(1-\beta)\sum_{i=1}^{n}W_{pop}(\mathrm{i})(V_i-V_j)\right) \quad (18)$$

3.4 算法描述

融合隐含信任度与项目关联度的矩阵分解推荐算法(TCRMF)步骤如下：

输入：用户-项目评分矩阵 R，用户-用户社交矩阵 S，潜在特征维数 d 和 k，正则化系数 λ_U 和 λ_V、超参数 α、β、σ 和 ρ，惩罚因子 γ，学习率 μ_1 和 μ_2，最大迭代次数 L_1 和 L_2。

输出：信任者的特征矩阵 B，被信任者的特征矩阵 W。用户特征阵 U 和项目特征矩阵 V。

begin
1. for $l_1<L_1$ do：
2. $B_i\leftarrow B_i-\mu_1\frac{\partial f_1}{\partial B_i}$，$W_j\leftarrow W_j-\mu_1\frac{\partial f_1}{\partial W_j}$；
3. $l_1=l_1+1$；
4. end for
5. 根据式(9)，计算用户 i 影响力 $Influence(i)$；
6. 根据式(10)，计算用户 i 和用户 j 的局部隐含信任度 $S_{local}(i,j)$；
7. 根据式(11)，计算用户 i 全局隐含信任度 $S_{global}(i)$；
8. 根据式(13)，计算项目 p 对项目 q 的相关度 $W_{associ}(p,q)$；
9. 根据式(14)，计算项目 p 流行度 $W_{pop}(p)$；
10. for $l_2<L_2$ do：
11. $U_i\leftarrow U_i-\mu_2\frac{\partial f_2}{\partial U_i}$，$V_j\leftarrow V_j-\mu_2\frac{\partial f_2}{\partial V_j}$；
12. $l_2=l_2+1$；
13. end for
end

TCRMF 算法的时间复杂度主要是根据目标函数 f_1、f_2 和相应偏导数的求解来计算的。目标函数 f_1 的时间复杂度是 $O(n\bar{t}d)$。目标函数 f_2 的时间复杂度是 $O(n\bar{r}k)$。其中，n 表示用户数量，d,k 分别表示算法的维度，$\bar{t}$ 表示用户平均信任关系数量，$\bar{r}$ 表示用户平均评分数量。由于信任和评分数据比较稀疏，$\bar{t}$、$\bar{r}$、d 和 k 的值都比较小，因此，目标函数 f_1 和 f_2 的计算是比较快的，且与总的用户数量 n 成线性正比关系。计算偏导数 $\frac{\partial f_1}{\partial B_i}$ 和 $\frac{\partial f_1}{\partial W_j}$ 的时间复杂度为 $O(n\bar{t}^2d)$，计算偏导数 $\frac{\partial f_2}{\partial U_i}$ 和 $\frac{\partial f_2}{\partial V_j}$ 的时间复杂度都为 $O(n\bar{r}k)$。由于 $\bar{t}$、$\bar{r}$、d 和 k 的值都比较小，因此偏导数的计算也是比较快的，且与总的用户数量 n 也成线性正比关系。所以目标函数和梯度计算的时间复杂度是 $O(n\bar{r}k+n\bar{t}d+n\bar{t}^2d)$。该算法的时间复杂度主要是受用户数量 n 的影响，因此该算法可以应用与大规模数据集。

4 实验结果和分析

本节主要介绍 TCRMF 算法与其他相关算法对推荐结果的影响。首先介绍实验采用的数据集和评价方法，然后设计 4 组实验从不同的角度对比分析本文算法的性能。

4.1 数据集

为了比较不同信息对推荐结果的影响，文本选择了两种公开数据集进行算法测试，Epinions(http://www.trustlet.org/epinions.html 和 Ciao(http://www.cse.msu.edu/~tangjili/trust.html)。这两个数据集都同时包含了用户评分信息和社交关系信息。两个数据集的详细统计特性如表 1 所示。

表 1　两个数据集统计特性

Table 1　Statistic of the two data sets

数据集	Epinions	Ciao
用户数量	31625	15375
项目数量	141368	106797
评分数量	583960	384086
社交关系数量	342033	151780

4.2 评价标准

为了评价推荐算法的性能，本文使用平均绝对误差(mean absolute error，MAE)和均方根误差(root mean squared error，RMSE)。这两种评价指标通过计算真实评分与预测评分之间的误差来衡量推荐结果的准确性，它们的值越小，推荐的精度就越高。两种评价指标的计算公式如式(19)和(20)所示。

$$MAE=\frac{1}{N}\sum_{(i,j)\in T}\left|R_{ij}-\hat{R}_{ij}\right| \quad (19)$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{(i,j)\in T}\left(R_{ij}-\hat{R}_{ij}\right)^2} \quad (20)$$

其中：N 表示测试集中评分的数量，T 表示测试集，R_{ij} 表示真实评分值，$\hat{R}_{ij}$ 表示预测评分值。

4.3 实验结果分析

为了验证 TCRMF 算法的预测准确率，将它与矩阵分解算法(MF)、Jamali 等人[22]提出基于信任传播的社会化推荐算法(SocialMF)和 Ma 等人[23]提出的基于社会正则化的推荐算法(SoReg)进行比较。

实验 1　不同特征向量维度下的实验分析

本实验比较了各种算法在 Epinions 和 Ciao 数据集的测试结果。对于所有基于矩阵分解的推荐算法，分别设置潜在特征维数为 5 和 10，并验证各种算法的精度。实验结果如表 2 和 3 所示。

a)SocialMF 和 SoReg 算法的 MAE 值和 RMSE 值都要小于 MF 算法，推荐精度有了一定的提高。这说明社交信息有助于提高传统矩阵分解推荐算法的预测精度。

b)本文的 TCRMF 算法相比于 SocialMF 和 SoReg 的社会化推荐算法，在 MAE 值和 RMSE 值上又有了进一步地提升。

这说明在社交信息基础上融合项目关联信息又可以进一步提高推荐算法的预测精度。

表 2　Epinions 数据集实验结果

Table 2　Experimental result of Epinions

	Metrics	MF	SocialMF	SoReg	TCRMF
k=5	MAE	0.942	0.893	0.871	**0.801**
	Improve	14.9%	10.3%	8.1%	
	RMSE	1.261	1.135	1.151	**1.023**
	Improve	18.9%	9.9%	11.1%	
k=10	MAE	0.902	0.863	0.851	**0.791**
	Improve	12.3%	8.3%	7.1%	
	RMSE	1.192	1.107	1.091	**1.013**
	Improve	15%	8.5%	7.1%	

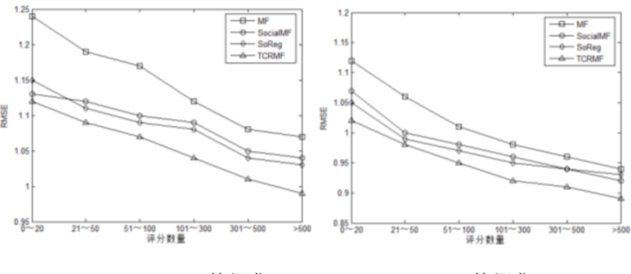
表 3　Ciao 数据集实验结果

Table 3　Experimental result of Ciao

	Metrics	MF	SocialMF	SoReg	TCRMF
k=5	MAE	0.905	0.861	0.856	**0.782**
	Improve	13.6%	9.2%	8.6%	
	RMSE	1.052	1.003	0.996	**0.931**
	Improve	11.5%	7.2%	6.5%	
k=10	MAE	0.872	0.834	0.823	**0.765**
	Improve	12.3%	8.3%	7%	
	RMSE	1.021	0.975	0.984	**0.903**
	Improve	11.6%	7.4%	8.2%	

实验 2　不同评分稀疏度的实验分析

为了测试不同评分稀疏度下的各种算法的效果。将不同评分数量划分为 0~20、21~50、51~100、101~300、301~500 和 500 以上一共 6 组。分别统计 Epinions 和 Ciao 两个数据集的 RMSE 值，实验结果如图 1 所示。



(a) Epinions 数据集　　(b) Ciao 数据集

图 1　不同评分稀疏度下 rmse 对比实验

Fig. 1　Comparsion on RMSE with different score sparsity

a)在不同评分稀疏度情况下，SocialMF、SoReg 和 TCRMF 算法的精度明显都要优于 MF 算法。这表明利用社交信息有助于提高算法的精度。

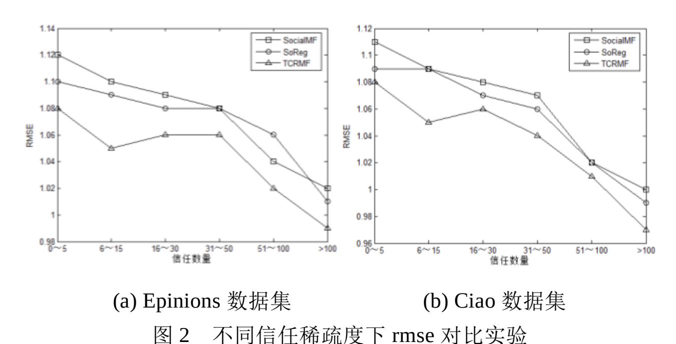
b)对于评分数据稠密(评分数量>300)的情况下，TCRMF 算法要明显优于其他基于社交信息的推荐算法。这表明随着评分数量的增加，本文算法能够有效的挖掘项目之间的关联关系，从而进一步提高推荐算法的精度。

实验 3　不同信任稀疏度的实验分析

为了测试不同社会关系下的各种算法的效果。将不同社会关系数量划分为 0~5、6~15、16~30、31~50、51~100 和 100 以上一共 6 组。本实验仅对基于社交信息的 SocialMF、SoReg 和 TCRMF 三种推荐算法进行对比实验，分别统计 Epinions 和 Ciao 数据集的 RMSE 值，实验结果如图 2 所示。

a)在不同信任稀疏度情况下，TCRMF 算法的精度都要优于 SocialMF、SoReg 算法。这表明在社交网络数据基础上融合项目的关联数据有助于提高算法的精度。

b)在信任数据稀疏(信任数据<15)的情况下，TCRMF 算法要明显优于其他基于社交信息的推荐算法。这表明对信任数据进行矩阵分解，通过综合局部隐含信任度和全局隐含信任度的推荐模型能有效的提高社会化推荐算法的精度。



(a) Epinions 数据集　　(b) Ciao 数据集

图 2　不同信任稀疏度下 rmse 对比实验

Fig. 2　Comparsion on RMSE with different trust sparsity

实验 4　不同项目关联关系的实验分析

为了评估不同项目关联度的计算方法对推荐算法精度的影响。首先根据皮尔森相似度计算项目间的关联度，并将项目的无向关联关系与 MF 结合得到改进的推荐算法 UCMF。然后根据本文提出的项目关联关系度量方法计算项目间的关联度，并将项目的有向关联关系与 MF 结合得到改进的推荐算法 DCMF。最后比较两种改进算法在 Epinions 和 Ciao 数据集的测试结果。分别设置潜在特征维数为 5 和 10，并计算两种算法的 RMSE 值。实验结果如表 4 所示。

表 4　关联关系对 rmse 的影响

Table 4　Impact of relations on RMSE

	Epinions k=5	Epinions k=10	Ciao k=5	Ciao k=10
UCMF	1.203	1.152	1.023	1.002
DCMF	1.196	1.108	1.016	0.995

a)相比于表 2 和 3 中 MF 算法，结合了项目间无向关联关系的 UCMF 算法的 RMSE 值都要小于 MF 算法。因为融合了项目的关联关系有助于提高算法的精度。

b)相比于 UCMF 算法，结合了项目间有向关联关系的 DCMF 算法的推荐精度都有了不同程度的提高。因为相比于无向的关联关系，本文方法能更好的反映了项目间真实的关联关系。

实验 5　参数对 RMSE 的影响

为了说明 TCRMF 算法中参数对推荐结果的影响，本文以 Epinions 数据集为例，从该数据集中抽取 80%作为训练集，比较各个参数下的推荐效果，实验结果如图 3 所示。

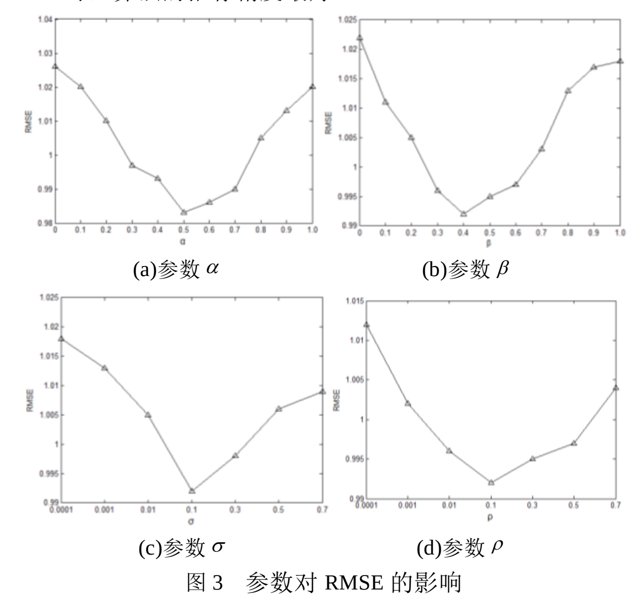
a)参数 α 控制用户的局部隐含信任度和全局隐含信任度对推荐算法的影响。由图 3(a)可知，当 α=1 时，算法只考虑了用户的局部隐含信任度信息；当 α=0 时，算法只考虑了用户的全局隐含信任度信息。当 $\alpha\in(0,1)$ 时，算法综合考虑了用户的局部隐含信任度和全局隐含信任度信息，随着 α 值的增加，算法的 RMSE 值先减小再增加，推荐精度先增加再减少。当 α=0.5 时，推荐精度最好。

b)参数 β 控制项目的相关度和流行度对推荐算法的影响。由图 3(b)可知，当 β=1 时，算法只考虑了项目的相关度信息；当 β=0 时，算法只考虑了项目的流行度信息。当 $\beta\in(0,1)$ 时，算法综合考虑了项目的相关度和流行度信息，随着 β 值的增加，推荐精度先增加再减少。当 β=0.4 时，推荐精度最好。

c)参数 σ 表示着用户的社会关系对推荐算法的影响。由

图 3(c)可知，随着 σ 值的增加，用户的社会关系在 TCRMF 算法所占比重逐渐增加，推荐精度先增加再减少。当 σ=0.1 时，算法取得较好的推荐精度。

d)参数 ρ 表示着项目的关联关系对推荐算法的影响。由图 3(d)可知，随着 ρ 值的增加，项目的关联关系在 TCRMF 算法所占比重逐渐增加，推荐精度也是先增加再减少。当 ρ=0.1 时，算法的推荐精度最好。



(a)参数 α　　(b)参数 β

(c)参数 σ　　(d)参数 ρ

图 3　参数对 RMSE 的影响

Fig. 3　Impact of parameters on RMSE

5　结束语

本文提出了一种融合隐含信任度和项目关联度的矩阵分解算法。利用矩阵分解模型将信任数据进行分解，得到用户的潜在被信任矩阵，并在此基础上引入用户的影响力，提出一种综合局部隐含信任度和全局的隐含信任度的推荐模型。该模型提高了对于信任数据稀疏时推荐系统的精度。为考虑项目间相互影响的差异性，在项目皮尔森相似度的基础上加入惩罚系数引入项目的相关度，同时从全局的角度考虑项目的流行度对项目关联度进行修正，提出了一种综合关联度推荐模型，该模型可以在不借助项目其他外部信息的基础上更好的反映了项目间的有向关系。实验结果表明，本文所提算法对于稀疏的信任数据要优于与现有的社会推荐算法，同时在社交信息基础上融合项目关联信息又可以进一步提高推荐算法的预测精度，取得更加明显的效果。在今后的工作中，笔者将把重点放在如何将地理位置等情景信息融入本文的推荐模型中来预测目标用户对项目的喜好，以期更好地提高该推荐算法的精度。

参考文献：

[1] 肖晓丽，钱娅丽，李旦江，等. 基于用户兴趣和社交信任的聚类推荐算法 [J]. 计算机应用，2016, 36(5): 1273-1278. (Xiao Xiaoli, Qian Yali, Li Danjiang, *et al*. Clustering recommendation algorithm based on user interest and social trust [J]. Journal of Computer Application, 2016, 36(5): 1273-1278.)

[2] Zhao Weidong, Wu Ran, Liu Haitao. Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target [J]. Information Processing and Management, 2016, 52 (5): 976-988.

[3] Tian Geng, Jing Liping. Recommending scientific articles using bi-relational graph-based iterative RWR [C]//Proc of the 7th ACM Conference on Recommender System. New York: ACM Press, 2013: 399-402.

[4] Sun Jianshan, Ma Jian, Liu Zhiying, *et al.* Leveraging content and connection for scientific article recommendation in social computing contexts [J]. Computer Journal, 2014, 57(9): 1331-1342.

[5] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems [J]. IEEE Computer, 2009, 42(8): 30-37.

[6] Mnih A, Salakhutdinov R. Probabilistic matrix factorization [C]// Proc of the 20th International Conference on Neural Information Processing Systems. New York: ACM Press, 2007: 1257-1264.

[7] Cremonesi P, Turrin R. Analysis of cold-start recommendations in IPTV system [C]// Proc of the 3th ACM Conference on Recommender System. New York: ACM Press, 2009: 233-236.

[8] Yang Xiwang, Guo Yang, Liu Yong, *et al.* A survey of collaborative filtering based social recommender systems [J]. Computer Communications, 2014, 41(5): 1-10.

[9] Massa P, Avesani P. Trust-aware Collaborative Filtering for Recommender Systems [C]//Proc of Federated International Conference on the Move to Meaningful Internet: CoopIS. Berlin: Springer Press, 2004: 492-508.

[10] 王兴茂，张兴明，邬江兴.基于一跳信任模型的协同过滤推荐算法 [J].通信学报, 2015, 36(6): 193-200. (Wang Xingmao, Zhang Xingming, Wu Jiangxing. Collaborative filtering recommendation algorithm based on one-jump turst model [J]. Journal on Communication, 2015, 36(6): 193-200.)

[11] 赵海燕，熊波，陈庆奎，等. 基于信任传播的概率矩阵分解算法 [J]. 小型微型计算机系统, 2016, 37(5): 895-901. (Zhao Haiyan, Xiong Bo, Chen Qingkui, *et al.* Based on the probability of trust propagation matrix decomposition algorithm [J]. Journal of Chinese Computer Systems, 2016, 37(5): 895-901.)

[12] 陈婷，朱青，周梦溪，等. 社交网络环境下基于信任的推荐算法 [J]. 软件学报, 2017, 28 (3): 721-731. (Chen Ting, Zhu Qing, Zhou Mengxi, *et al.* Trust-based recommendation algorithm in social network [J]. Journal of Software, 2017, 28 (3): 721-731.)

[13] 郭磊，马军，陈竹敏. 一种信任关系强度敏感的社会化推荐算法 [J]. 计算机研究与发展, 2015, 50(9): 1805-1813. (Guo Lei, Ma Jun, Chen Zhumin. Trust strength aware social recommendation method [J]. Journal of Computer Research and Development, 2015, 50 (9): 1805-1813.)

[14] Wu Le, Chen Enhong, Liu Qi, *et al.* Leveraging tagging for neighborhood-aware probabilistic matrix factorization [C]// Proc of the 21st ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2012: 1854-1858.

[15] 杨兴耀，于炯，吐尔根•依布拉音，等. 结合用户和项目预测的协同过滤模型 [J]. 计算机应用, 2013, 33 (12): 3354-3358. (Yang Xingyao, Yu Jiong, Turgun Ibrahim, *et al.* Collaborative filtering model combining users' and items' predictions [J]. Journal of Computer Applications, 2013, 33(12): 3354-3358.)

[16] 于美琪，邝砾，禹斌，等. 融合社会网络和项目特征的移动应用推荐 [J]. 小型微型计算机系统, 2017, 38(2): 310-313. (Yu Meiqi, Kuang Li, Guo Bin, *et al.* App recommendation combing item features and social networks [J]. Journal of Chinese Computer Systems, 2017, 38 (2): 310-313.)

[17] 孙金刚，艾丽蓉. 基于项目属性和云填充的协同过滤推荐算法 [J]. 计算机应用, 2012, 32(3): 658-660. (Sun Jingang, Ai Lirong. Collaboratvie filtering recommendation algorithm based on item

attribute and cloud model filling [J]. Journal of Computer Application, 2012, 32 (3): 658-660.)

[18] Yang Bo, Lei Yu, Liu Jiming, *et al*. Social collaborative filtering by trust [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2017, 39 (8): 1633-1647.

[19] 张燕平，张顺，钱付兰，等. 一种局部和全局用户影响力相结合的社交推荐算法 [J]. 南京大学学报：自然科学版 , 2015, 51(4): 858-865. (Zhang Yanping, Zhang Shun, Qian Fulan, *et al*. Local and global user influence combined social recommendation algorithms [J]. Journal of Nanjing University: Natural Sciences, 2015, 51(4): 858-865.)

[20] Tang Jiliang, Hu Xia, Liu Huan. Social recommendation: a review [J]. Social Network Analysis and Mining. 2013, 3(4): 1113-1133.

[21] 王锦坤，姜元春，孙见山. 考虑用户活跃度和项目流行度的基于项目最近邻的协同过滤算法 [J]. 计算机科学，2016，43(12)：158-162. (Wang Jinkun, Jiang Yuanchun, Sun Jianshan, *et al*. Item-based collaborative filtering algorithm integrating user activity and item popularity [J]. Computer Science, 2016, 43(12): 158-162.)

[22] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks [C]// Proc of the 4th lnternational Conference on Recommender Systems. New York: ACM Press, 2010: 135-142.

[23] Ma Hao, Zhou Dengyong, Liu Chao, *et al*. Recommender systems with social regularization [C]// Proc of the 25th lnternational Conference on Web Search and Data Mining. New York: ACM Press, 2011: 287−296.